JOURNAL OF MULTIVARIATE ANALYSIS **62**, 121–136 (1997)
ARTICLE NO. MV971677

On Estimating the Dimensionality in Canonical Correlation Analysis

Brenda K. Gunderson and Robb J. Muirhead*

University of Michigan

In canonical correlation analysis the number of nonzero population correlation coefficients is called the dimensionality. Asymptotic distributions of the dimensionalities estimated by Mallows's criterion and Akaike's criterion are given for nonnormal multivariate populations with finite fourth moments. These distributions

totic distribution of the corresponding estimator is derived for multivariate normal populations. It is shown that this estimator is not consistent, but that a simple modification yields consistency. An overall comparison of the various estimation methods is conducted through simulation studies. © 1997 Academic Press

1. INTRODUCTION

Canonical correlation analysis focuses on the relationship between two sets of random variables, X $(p \times 1)$ and Y $(q \times 1)$, $p \leqslant q$. Put

$$W = \begin{pmatrix} X \\ Y \end{pmatrix},$$

and let $\rho_1^2, ..., \rho_p^2$ $(1 > \rho_1^2 \geqslant \cdots \geqslant \rho_p^2 \geqslant 0)$ be the eigenvalues of the matrix $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$, where

$$\operatorname{cov}(W) = \operatorname{cov} \begin{pmatrix} X \\ Y \end{pmatrix} = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Received July 24, 1996.

Key words and phrases: Akaike's information criterion, canonical correlation coefficient, dimensionality, elliptical distribution, kurtosis, likelihood, Mallows's criterion.

* Work supported by the National Science Foundation and a Research Fellowship at the Australian National University.

0047-259X/97 $25.00

Their positive square roots $\rho_1, ..., \rho_p$ are the population canonical correlation coefficients. Let

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

be the sample covariance matrix formed from a random sample of size $N = n+1$ from the distribution of W. The sample canonical correlation coefficients $r_1, ..., r_p$ (where $1 > r_1 \geqslant \cdots \geqslant r_p > 0$) are the positive square roots of the eigenvalues of $S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}$.

The *dimensionality* in canonical correlation analysis is K, the number of *nonzero* population canonical correlation coefficients. This corresponds to the number of pairs of canonical variables that are deemed useful in a practical setting. In this paper we investigate the behavior of some estimators of K based on "model selection" procedures. Two of these have been studied by Fijikoshi and Veitch (1979) and Fujikoshi (1985) under normal sampling; we extend their results in Sections 2 and 3 to the nonnormal setting. A new estimator based on a marginal likelihood function for K is also proposed in Section 4 and its asymptotic distribution is derived. The estimators are compared through simulation studies presented in Section 5.

2. AKAIKE'S INFORMATION CRITERIA

Based on Akaike's information criterion for model choice, Fujikoshi and Veitch (1979), assuming normal sampling, proposed the estimator $\hat{K}_A$ of K given by

$$\hat{K}_A = k \qquad \text{when} \quad A_k = \min(A_0, ..., A_p),$$

where

$$A_k = -N \sum_{i=k+1}^{p} \log(1 - r_i^2) - 2(p-k)(q-k), \qquad k = 0, ..., p-1,$$

with $A_p \equiv 0$. Fujikoshi (1985) studied asymptotic properties of this estimator under normal sampling. Here we give the asymptotic distribution of $\hat{K}_A$ when the population being sampled is elliptical and when $K = k_0$ (so that $\rho_1 > \cdots > \rho_{k_0} > \rho_{k_0+1} = \cdots = \rho_p = 0$). The result is given in the following theorem. This is a special case of a more general theorem, holding for general nonnormal population with finite fourth moments, which appears (together with a sketch of its proof) in the Appendix as Theorem A1. In the following theorem 3κ is the kurtosis of each of the marginal distributions formed from a $(p+q)$-variate elliptical distribution.

THEOREM 1. *Let r_i $(i=1, ..., p)$ be the sample canonical correlation coefficients formed from a random sample of size $N=n+1$ drawn from a $(p+q)$-variate elliptical distribution with kurtosis parameter κ. Suppose $K=k_0$. Then*

$$\lim_{n \to \infty} P(\hat{K}_A = k) \equiv P(k \mid k_0),$$

where

$$P(k \mid k_0) = \begin{cases} 0, & 0 \leqslant k < k_0, \\ P\left(\sum_{i=1}^{p_{mk}} U_i > \frac{2p_{mk}}{1+\kappa},\ m=k_0, ..., k-1\right), & \\ \quad \times P\left(\sum_{j=1}^{p_{km}} V_j \leqslant \frac{2p_{km}}{1+\kappa},\ m=k+1, ..., p\right), & k_0 \leqslant k \leqslant p, \end{cases}$$

with the U_i and V_j all being independent χ_1^2 variables, and $p_{ij} \equiv (p-i)(q-i) - (p-j)(q-j)$.

Obviously the probabilities in Theorem 1 can also be expressed as joint probabilities involving correlated χ^2 random variables. From Theorem 1 we see that the asymptotic probability of correctly assessing the dimensionality is

$$P(k_0 \mid k_0) = P\left(\sum_{j=1}^{p_{km}} V_j \leqslant \frac{2p_{k_0 m}}{1+\kappa},\ m=k_0+1, ..., p\right),$$

so that the estimator $\hat{K}_A$ is not consistent if $k_0 < p$. If we take $k=k_0=p-1$ then we have

$$P(k_0 \mid k_0) = P\left(\chi^2_{q-p+1} \leqslant \frac{2(q-p+1)}{1+\kappa}\right).$$

These results all agree with those of Fujikoshi (1985) under normal sampling (when $\kappa = 0$).

Consistency adjustment. Lack of consistency is common with Akaike-like estimators, but the criterion can often be easily modified to yield consistent estimators. Here an appropriate modified criterion (based on a procedure due to Schwarz, 1978) is

$$S_k = -N \sum_{i=k+1}^{p} \log(1-r_i^2) - (\log n)\{(p-k)(q-k)\},$$

where $S_p = 0$. It may be shown (see Gunderson, 1989) that the estimator $\hat{K}_S$ of K defined by

$$\hat{K}_S = k \qquad \text{when} \quad S_k = \min(S_0, ..., S_p),$$

is a consistent estimator of K, provided the population being sampled has finite fourth moments.

Kurtosis adjustment. We see from the preceding results that if $\hat{K}_A$ is used to estimate the dimensionality K when the underlying population is long-tailed elliptical, the asymptotic probability of correctly assessing K decreases. For elliptical distributions we propose a modified criterion to adjust for nonzero kurtosis, namely

$$A_k^* = \left(\frac{1}{1+\kappa}\right)\left\{-N \log \prod_{i=k+1}^{p} (1-r_i^2)\right\} - 2(p-k)(q-k),$$

and define a "kurtosis adjusted" estimator $\hat{K}_A^*$ given by

$$\hat{K}_A^* = k \qquad \text{when} \quad A_k^* = \min(A_0^*, ..., A_p^*).$$

Then the asymptotic probability of correctly assessing the dimensionality when $k = k_0 = p - 1$ becomes

$$\lim_{n \to \infty} P(\hat{K}_A^* = k_0) = P(k_0 \mid k_0) = P\{\chi^2_{q-p+1} \leqslant 2(q-p+1)\},$$

the same as it is for normal sampling.

A simulation study (see Section 5) showed that the adjusted estimator $\hat{K}_A^*$ generally performs better for long-tailed elliptical distributions than the normal-based estimator $\hat{K}_A$. (Of course, if $\hat{\kappa}$ is a consistent estimator of κ, the limiting distribution is unchanged if we replace κ by $\hat{\kappa}$.)

3. MALLOW'S CRITERION

Based on Mallow's C_p statistic, Fujikoshi and Veitch (1979) proposed the estimator $\hat{K}_C$ of K given by

$$\hat{K}_C = k \qquad \text{when} \quad C_k = \min(C_0, ..., C_p),$$

where

$$C_k = N \sum_{i=k+1}^{p} \left(\frac{r_i^2}{1-r_i^2}\right) - 2(p-k)(q-k), \qquad k=0, ..., p-1,$$

with $C_p \equiv 0$. Fujikoshi (1985) showed that, in the case of normal sampling, the asymptotic distribution of $\hat{K}_C$ is exactly the same as that of $\hat{K}_A$, the estimator obtained using Akaike's criteron. It is shown in Gunderson (1989) that this remains true for general nonnormal distributions with finite fourth moments. In particular, if the parent population is elliptical, the asymptotic distribution of $\hat{K}_C$ is given by Theorem 1. As in the case of Akaike's criterion, we propose a modified criterion to adjust for nonzero kurtosis, namely

$$C_k^* = \left(\frac{1}{1+\kappa}\right) N \sum_{i=k+1}^{p} \left(\frac{r_i^2}{1-r_i^2}\right) - 2(p-k)(q-k), \tag{2.6}$$

with the "kurtosis adjusted" estimator being $\hat{K}_C^*$, given by

$$\hat{K}_C^* = k \qquad \text{when} \quad C_k^* = \min(C_0^*, ..., C_p^*),$$

Again, substituting a consistent estimate of $\hat{\kappa}$ for κ in the modified criterion C_k^* would define a useful criterion for practical applications.

4. MARGINAL LIKELIHOOD CRITERION

In the case of normal sampling, Glynn and Muirhead (1978) (see Muirhead, 1982, Theorem 11.3.6) gave an asymptotic representation (for large n) for the joint probability density function of $r_1^2, ..., r_p^2$ when the population canonical correlation coefficients satisfy $\rho_1 > \cdots > \rho_k > \rho_{k+1} \cdots = \rho_p = 0$ (i.e., the dimensionality is $K=k$). This then defines a likelihood function for K and $\rho_1, ..., \rho_K$ which, when $\rho_1, ..., \rho_K$ are replaced by their maximum likelihood estimators $r_1, ..., r_K$, yields a "marginal likelihood function" for K. Full details may be found in Gunderson (1989). Maximizing this marginal likelihood function for K leads to the "maximum marginal likelihood estimator" $\hat{K}_M$ of K defined by

$$\hat{K}_M = k \qquad \text{when} \quad l_k = \max(l_0, ..., l_p),$$

where l_k is the marginal log-likelihood for K given by

$$l_k = \frac{(p+q+1-n)}{2} \sum_{i=1}^{k} \log(1-r_i^2) + \frac{(k-q)}{2} \sum_{i=1}^{k} \log(r_i^2)$$
$$- \sum_{i<j}^{k} \log(r_i^2 - r_j^2) - \frac{1}{2} \sum_{i=1}^{k} \sum_{j=k+1}^{p} \log(r_i^2 - r_j^2) - \frac{k(p+q-k-1)}{2} \log\left(\frac{n}{2}\right)$$
$$- k \log(2\pi) + \sum_{i=1}^{k} \left\{ \log \Gamma\left(\frac{p-i+1}{2}\right) + \log \Gamma\left(\frac{q-i+1}{2}\right) \right\},$$
$$k = 1, 2, ..., p$$

and $l_0 \equiv 0$.

The following theorem, whose proof is sketched in the Appendix, gives the asymptotic distribution of $\hat{K}_M$ under normal sampling when the true dimensionality is $K = k_0$.

THEOREM 2. *Suppose that $K = k_0$. Then*

$$\lim_{n \to \infty} P(\hat{K}_M = k) \equiv P(k \mid k_0),$$

where

$$P(k \mid k_0) = \begin{cases} 0, & 0 \leqslant k < k_0, \\ P\{Z(k, m, k_0) > 2q_{km},\ m = k_0, ..., k-1,\ \textit{and} & \\ \qquad Z(m, k, k_0) \leqslant 2q_{mk},\ m = k+1, ..., p\}, & k_0 \leqslant k \leqslant p, \end{cases}$$

with

$$Z(k, m, k_0) = \sum_{i=m+1}^{k} x_i + (k-q) \sum_{i=m+1}^{k} \log(x_i) + (k-m) \sum_{i=k_0+1}^{m} \log(x_i)$$
$$- 2 \sum_{\substack{i=m+1}}^{k} \sum_{\substack{j=m+1 \\ j>i}}^{k} \log(x_i - x_j) - \sum_{i=m+1}^{k} \sum_{j=k+1}^{p} \log(x_i - x_j)$$
$$- \sum_{i=k_0+1}^{m} \sum_{j=m+1}^{k} \log(x_i - x_j),$$

$$q_{km} = \sum_{i=m+1}^{k} \left\{ \log(2\pi) - \log \Gamma\left(\frac{p-i+1}{2}\right) - \log \Gamma\left(\frac{q-i+1}{2}\right) - \frac{(p+q-k-m-1)}{2} \log(2) \right\}$$

and the x_i's $(i=k_0+1, ..., p)$ being random variables having the same distribution as the eigenvalues of a $(p-k_0)$ by $(p-k_0)$ matrix with a $W_{p-k_0}(q-k_0, I_{p-k_0})$ Wishart distribution.

From Theorem 2 we see that $\hat{K}_M$ is not consistent, as asymptotically there is a positive probability of overestimating the true dimensionality. As with Akaike's criterion, a slight modification yields a consistent estimator. Let $l_k^* = l_k - \frac{1}{2}\log n$ (where l_k is the marginal log-likelihood function for K), and consider the estimator $\hat{K}_{MC}$ given by

$$\hat{K}_{MC}=k \qquad \text{when} \quad l_k^* = \max(l_0^*, ..., l_p^*).$$

It may be shown (see Gunderson, 1989) that

$$\lim_{n\to\infty} P(\hat{K}_{MC}=k_0)=P(k_0 \mid k_0)=1,$$

so that $\hat{K}_{MC}$ is a consistent estimator of K.

5. SIMULATION STUDIES

The various estimators of K were compared in simulations using normal and elliptical $t(5)$ distributions. Three sample sizes (N = 50, 100, 200) were used, with one choice of p and q, namely $p=q=4$. Table I shows eight values of $P=\text{diag}(\rho_1, \rho_2, \rho_3, \rho_4)$ which were used in the simulation study. For normal sampling, the study consisted of generating 500 values of a Wishart matrix S, $S \sim W_8(n, \Sigma)$, with $n=N-1$ and where Σ has the form

$$\Sigma = \begin{pmatrix} I_p & P \\ P' & I_q \end{pmatrix}. \tag{5.1}$$

For elliptical $t(5)$ sampling, the study consisted of generating 500 samples of size N of an 8×1 random vector W having an elliptical t distribution on 5 degrees of freedom and parameters $\mu=0$ and scale matrix $V=\frac{3}{5}\Sigma$, with Σ of the form (5.1).

TABLE I

P Matrices Used in the Simulation Studies

$10\rho_1$	8	3	9	5	9	5	9	5
$10\rho_2$	0	0	8	3	8	4	8	4
$10\rho_3$	0	0	0	0	7	3	7	3
$10\rho_4$	0	0	0	0	0	0	6	2
K	1	1	2	2	3	3	4	4

TABLE II

The Percentage of Correct and Incorrect Estimation of the Dimension under Normal Sampling

P	K	N	Percentage	$\hat{K}_M$	$\hat{K}_{MC}$	$\hat{K}_A$	$\hat{K}_C$	$\hat{K}_S$
diag(.8, 0, 0, 0)	1	50	Under	00.0	00.2	00.0	00.0	00.2
			Correct	85.8	99.2	79.0	70.8	99.2
		100	Under	00.0	00.0	00.0	00.0	00.0
			Correct	84.2	99.2	82.2	80.2	99.8
		200	Under	00.0	00.0	00.0	00.0	00.0
			Correct	82.8	99.4	81.4	79.6	100.
diag(.3, 0, 0, 0)	1	50	Under	74.0	95.6	50.6	37.4	96.8
			Correct	17.2	04.4	40.0	48.6	03.2
		100	Under	45.8	81.8	32.6	27.0	96.4
			Correct	38.8	17.8	56.6	60.2	03.6
		200	Under	09.8	43.2	05.6	05.2	85.8
			Correct	68.2	56.2	79.4	77.4	14.2
diag(.9, .8, 0, 0)	2	50	Under	00.0	00.0	00.0	00.0	00.0
			Correct	79.6	98.6	82.0	79.2	98.0
		100	Under	00.0	00.0	00.0	00.0	00.0
			Correct	77.0	99.4	83.0	81.8	99.4
		200	Under	00.0	00.0	00.0	00.0	00.0
			Correct	77.0	99.6	84.6	84.4	100.
diag(.5, .3, 0, 0)	2	50	Under	65.0	94.2	59.0	50.8	96.8
			Correct	21.0	05.6	34.6	41.4	03.2
		100	Under	31.6	76.4	27.2	24.0	93.6
			Correct	48.8	23.2	63.6	65.2	06.4
		200	Under	05.2	30.2	04.0	03.2	69.4
			Correct	71.4	69.0	84.2	84.6	30.6
diag(.9, .8, .7, 0)	3	50	Under	00.0	00.4	00.0	00.0	00.2
			Correct	77.0	98.4	82.2	81.6	93.4
		100	Under	00.0	00.0	00.0	00.0	00.0
			Correct	77.0	99.0	82.8	82.4	96.4
		200	Under	00.0	00.0	00.0	00.0	00.0
			Correct	76.2	99.4	81.8	81.6	98.0
diag(.5, .4, .3, 0)	3	50	Under	58.4	96.4	66.6	61.8	98.4
			Correct	26.4	03.6	26.8	31.0	01.4
		100	Under	18.6	71.6	25.0	23.8	88.2
			Correct	61.6	27.8	62.8	64.0	11.6
		200	Under	01.8	17.8	02.6	02.4	39.8
			Correct	74.0	81.6	81.8	81.8	58.8

TABLE II—*Continued*

P	K	N	Percentage	$\hat{K}_M$	$\hat{K}_{MC}$	$\hat{K}_A$	$\hat{K}_C$	$\hat{K}_S$
diag(.9, .8, .7, .6)	4	50	Under	00.0	02.2	00.0	00.0	00.2
			Correct	100.	97.8	100.	100.	99.8
		100	Under	00.0	00.0	00.0	00.0	00.0
			Correct	100.	100.	100.	100.	100.
		200	Under	00.0	00.0	00.0	00.0	00.0
			Correct	100.	100.	100.	100.	100.
diag(.5, .4, .3, .2)	4	50	Under	68.6	98.0	86.2	83.6	99.2
			Correct	31.4	02.0	13.8	16.4	00.8
		100	Under	32.8	87.0	48.2	47.8	91.8
			Correct	67.2	13.0	51.8	52.2	08.2
		200	Under	08.8	51.6	13.4	13.4	49.0
			Correct	91.2	48.4	86.6	86.6	51.0

Note. Percent of overestimation can be obtained by difference from 100.

For each estimation method the number of times each possible value (0, 1, 2, 3, 4) of the dimension appeared as an estimate was recorded, and the percentage of underestimation and overestimation was then obtained.

Results for normal sampling. Table II presents the percentage of correct and incorrect estimates of the dimensionality for the eight matrices in Table I. As expected, the percentages for Akaike's method and Mallow's method are very close, especially for large sample sizes. Although the asymptotic distributions of $\hat{K}_A$ and $\hat{K}_C$ depend only on k_0, p, and q, the speed of convergence depends highly on the values of the population canonical correlation correlations, and especially on the value of $\rho_{\min}$.

In situations where either or both N or $\rho_{\min}$ are small, the estimators $\hat{K}_A$ and $\hat{K}_C$ perform better than the consistent estimators $\hat{K}_S$ and $\hat{K}_{MC}$, with $\hat{K}_S$ usually outperformed by $\hat{K}_{MC}$. When N is large and $\rho_{\min}$ is appreciable, the consistent estimators perform best.

To avoid overestimation the consistent criteria should be used, but sometimes at the cost of underestimation. To avoid underestimation, Akaike's and Mallow's criteria are recommended. The general level of correct estimation, as expected, decreases with decreasing sample size and decreasing nonzero population canonical correlation coefficients.

Results for elliptical $t(5)$ sampling. Table III gives the percentages of correct and incorrect estimates of the dimensionality for the eight matrices in Table I. The estimators $\hat{K}_A^*$ and $\hat{K}_C^*$ adjusted for nonzero kurtosis perform better than their normal-based counterparts $\hat{K}_A$ and $\hat{K}_C$ when the sample size N is large and $\rho_{\min}$ is appreciable, while for small values of $\rho_{\min}$ the

TABLE III

The Percentage of Correct and Incorrect Estimation of the Dimension under Elliptical $t(5)$ Sampling

P	K	N	Percantage	$\hat{K}_M$	$\hat{K}_{MC}$	$\hat{K}_A$	$\hat{K}_A^*$	$\hat{K}_C$	$\hat{K}_C^*$	$\hat{K}_S$
diag(.8, 0, 0, 0)	1	50	Under	00.0	00.0	00.0	07.8	00.0	00.2	00.0
			Correct	51.8	83.8	43.0	90.6	37.2	93.8	87.0
		100	Under	00.0	00.0	00.0	00.0	00.0	00.0	00.0
			Correct	42.6	82.2	41.0	96.4	38.2	94.0	90.4
		200	Under	00.0	00.0	00.0	00.0	00.0	00.0	00.0
			Correct	36.2	77.4	36.6	93.8	35.4	95.0	92.0
diag(.3, 0, 0, 0)	1	50	Under	37.6	68.8	17.6	95.2	11.2	84.2	76.4
			Correct	32.6	25.2	47.6	04.8	47.0	14.8	20.6
		100	Under	13.6	44.4	10.0	89.8	08.0	78.8	74.4
			Correct	38.4	47.2	45.4	09.8	44.6	19.8	22.8
		200	Under	02.0	18.8	01.6	70.2	01.4	57.2	54.8
			Correct	38.2	64.8	44.8	28.8	42.2	41.2	42.8
diag(.9, .8, 0, 0)	2	50	Under	00.0	00.0	00.0	02.8	00.0	00.6	00.2
			Correct	57.0	88.8	56.6	94.4	52.6	93.4	87.2
		100	Under	00.0	00.0	00.0	07.8	00.0	00.0	00.0
			Correct	51.0	87.8	55.2	94.4	53.4	93.6	89.8
		200	Under	00.0	00.0	00.0	00.0	00.0	00.0	00.0
			Correct	44.4	86.2	49.0	95.0	48.2	93.8	92.4
diag(.5, .3, 0, 0)	2	50	Under	41.6	82.6	35.6	98.8	26.8	94.8	88.2
			Correct	31.6	14.2	44.0	01.2	49.8	05.0	09.8
		100	Under	14.4	55.0	13.6	92.8	10.6	86.8	79.4
			Correct	42.8	40.4	57.2	06.6	57.6	12.2	18.6
		200	Under	02.2	22.4	02.0	63.2	01.8	58.4	52.8
			Correct	44.6	69.8	57.0	36.2	55.2	40.8	45.4
diag(.9, .8, .7, 0)	3	50	Under	00.4	02.2	00.0	08.2	00.0	04.2	01.6
			Correct	65.8	93.2	73.0	87.6	72.8	91.2	88.0
		100	Under	00.0	00.2	00.0	00.2	00.0	00.2	00.2
			Correct	61.0	93.2	70.0	92.6	69.8	92.0	88.8
		200	Under	00.0	00.0	00.0	00.0	00.0	00.0	00.0
			Correct	62.2	92.4	68.4	90.6	68.4	90.0	88.6
diag(.5, .4, .3, 0)	3	50	Under	52.8	89.2	59.4	99.0	55.6	97.8	93.2
			Correct	29.6	10.0	32.0	01.0	35.4	02.0	05.4
		100	Under	19.4	68.4	25.2	93.0	24.0	88.4	80.2
			Correct	50.8	30.0	56.0	06.4	57.2	10.8	18.4
		200	Under	03.2	25.2	05.0	52.9	05.0	49.8	40.4
			Correct	59.0	69.8	70.0	43.4	69.6	46.0	53.2

TABLE III—*Continued*

P	K	N	Percantage	$\hat{K}_M$	$\hat{K}_{MC}$	$\hat{K}_A$	$\hat{K}_A^*$	$\hat{K}_C$	$\hat{K}_C^*$	$\hat{K}_S$
diag(.9, .8, .7, .6)	4	50	Under	00.6	06.0	01.0	07.4	01.0	06.6	03.4
			Correct	99.4	94.0	99.0	92.6	99.0	93.4	96.6
		100	Under	00.0	00.6	00.0	00.6	00.0	00.6	00.6
			Correct	100.	99.4	100.	99.4	100.	99.4	99.4
		200	Under	00.0	00.0	00.0	00.0	00.0	00.0	00.0
			Correct	100.	100.	100.	100.	100.	100.	100.
diag(.5, .4, .3, .2)	4	50	Under	73.4	97.0	84.6	99.8	84.0	99.8	96.6
			Correct	26.6	03.0	15.4	00.2	16.0	00.2	03.4
		100	Under	47.6	90.2	62.2	96.4	61.8	94.0	90.8
			Correct	52.4	09.8	37.8	03.6	38.2	06.0	09.2
		200	Under	17.0	65.4	26.8	69.0	26.6	67.6	62.2
			Correct	83.0	34.6	73.2	31.0	73.4	32.4	37.8

Note. Percent of overestimation can be obtained by difference from 100.

normal-based estimators generally outperform their modified counterparts. Out of the two modified estimators, $\hat{K}_C^*$ performs a little better overall than $\hat{K}_A^*$. Of the two consistent estimators $\hat{K}_S$ and $\hat{K}_{MC}$, the latter performs better for small values of $\rho_{\min}$, while the former is better for moderate to large values of $\rho_{\min}$.

To avoid overestimation of the dimensionality the modified estimators $\hat{K}_A^*$ and $\hat{K}_C^*$ should be used. A consequence of this, however, is an increase in the percentage of underestimation. The consistent estimators also have a tendency to underestimate the dimensionality, but not as extremely as the modified estimators.

APPENDIX: PROOFS

Preliminaries. First, some preliminaries which are needed in the sequel. Because of invariance considerations, we can assume without loss of generality that the population covariance matrix has the form (5.1). Assuming fourth moments are finite, the fourth order cumulants of the distribution of the $(p+q)\times 1$ random vector W are given by (see, for example, Muirhead and Waternaux, 1980)

$$\kappa_{1111}^{ijkl} = E(W_i W_j W_k W_l) - \sigma_{ij}\sigma_{kl} - \sigma_{ik}\sigma_{jl} - \sigma_{il}\sigma_{jk}.$$

For the particular covariance structure in (5.1) the kurtosis of the ith component is

$$\kappa_4^i = E(W_i^4) - 3$$

and the bivariate fourth-order cumulants are

$$\kappa_{22}^{ij} = \operatorname{cov}(W_i^2, W_j^2) = E(W_i^2 W_j^2) - 1 \qquad (j \neq i + p),$$
$$\kappa_{22}^{i,\, i+p} = E(W_i^2 W_{i+p}^2) - 1 - 2p_i^2.$$

In elliptical distributions all fourth-order cumulants are determined by one parameter κ as

$$\kappa_{1111}^{ijkl} = \kappa(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}),$$

where κ characterizes the kurtosis of the distribution. For the particular covariance structure (5.1) we have

$$\kappa_{1111}^{i,\, i+p,\, j,\, j+p} = \rho_i \rho_j \kappa, \qquad \kappa_{31}^{i,\, i+p} = \kappa_{13}^{i,\, i+p} = 3\rho_i \kappa$$
$$\kappa_{22}^{i,\, i+p} = (1 + 2\rho_i^2)\kappa, \qquad \kappa_{22}^{ij} = \kappa \qquad (j \neq i,\, i+p),$$

and all other cumulants of order 4 are zero.

Proof of Theorem 1. We sketch the proof of a general result (stated later as Theorem A1) about the asymptotic distribution of $\hat{K}_A$, of which Theorem 1 is a special case. Assume that the $(p+q)$-variate distribution being sampled has finite fourth moments and that the true dimensionality is $K = k_0$. Since r_i converges to ρ_i in probability $(i = 1, ..., p)$ we have

$$\frac{1}{N} A_k \rightarrow -\log \prod_{i=k+1}^{k_0} (1 - \rho_i^2) \qquad (0 \leqslant k < k_0),$$
$$\frac{1}{N} A_k \rightarrow 0 \qquad (k_0 \leqslant k \leqslant p),$$

in probability, as $N \rightarrow \infty$. This implies that

$$\begin{aligned} &\lim_{n \rightarrow \infty} P(\hat{K}_A = k) = 0, && (0 \leqslant k < k_0), \\ &\lim_{n \rightarrow \infty} P(\hat{K}_A = k) = \lim_{n \rightarrow \infty} P(A_k \leqslant A_m,\ m = k_0, ..., p), && (k_0 \leqslant k \leqslant p). \end{aligned} \qquad \text{(A.1)}$$

We thus need only be concerned with the asymptotic behavior of $\hat{K}_A$ for $k_0 \leqslant k \leqslant p$. Using an expansion for $-N \sum_{i=k+1}^{p} \log(1-r_i^2)$ in Muirhead and Waternaux (1980), A_k can be expanded for large N as

$$A_k = \sum_{i=k+1}^{p} \sum_{j=k+1}^{q} z_{i,p+j}^2 - 2(p-k)(q-k) + O_p(N^{-1/2}),$$
$$k_0 \leqslant k \leqslant p, \tag{A.2}$$

where $z_{i,p+j} \equiv N^{1/2} s_{i,p+j}$, with $s_{i,p+j}$ being the $(i, p+j)$ element of the sample covariance matrix S. Define the $(p-k_0)(q-k_0) \times 1$ vector z by

$$\begin{aligned} z = (&z_{k_0+1,p+k_0+1}, ..., z_{k_0+1,p+q}; ...; z_{k,p+k_0+1}, ..., z_{k,p+q}; \\ &z_{k+1,p+k_0+1}, ..., z_{k+1,p+k}; ...; z_{p,p+k_0+1}, ..., z_{p,p+k}; \\ &z_{k+1,p+k+1}, ..., z_{k+1,p+q}; ..., z_{p,p+k+1}, ..., z_{p,p+q})'. \end{aligned}$$

By the multivariate central limit theorem the asymptotic distribution of z is normal with mean vector 0 and covariance matrix Ω which can be expressed in terms of the fourth-order cumulants of the parent population, namely,

$$\begin{aligned} \operatorname{var}(z_{i,p+j}) &= 1 + \kappa_{22}^{i,p+j} \\ \operatorname{cov}(z_{i,p+j}, z_{i,p+l}) &= \kappa_{211}^{i,p+j,p+l} \qquad (j \neq l) \\ \operatorname{cov}(z_{i,p+j}, z_{l,p+j}) &= \kappa_{211}^{p+j,i,l} \qquad (i \neq l) \\ \operatorname{cov}(z_{i,p+j}, z_{l,p+m}) &= \kappa_{1111}^{i,p+j,l,p+m} \qquad (i \neq l,\ j \neq m). \end{aligned}$$

By the spectral decomposition theorem there exists an orthogonal matrix H which diagonalizes Ω, the diagonal elements ω_i being the eigenvalues of Ω. If we put $v = Hz$, we have

$$v'v = \sum_{i=k_0+1}^{p} \sum_{j=k_0+1}^{q} z_{i,p+j}^2$$

and

$$v = Hz \to \mathcal{N}(0, H\Omega H')$$

in distribution, as $n \to \infty$. Thus the elements v_i of v are asymptotically independent normal with mean 0 and variance ω_i. Substituting (A.2) into (A.1) and using the limiting distribution of the variables $z_{i,p+j}$, we obtain the result given in the following theorem.

THEOREM A.1. *Let r_i $(i=1, ..., p)$ be the sample canonical correlation coefficients formed from a random sample of size $N=n+1$ drawn from a $(p+q)$-variate distribution with finite fourth moments, and assume the true dimensionality is $K=k_0$. Then*

$$\lim_{n \to \infty} P(\hat{K}_A = k) \equiv P(k \mid k_0),$$

where

$$P(k \mid k_0) = \begin{cases} 0, & 0 \leqslant k < k_0, \\ P\left(\sum_{i=1}^{p_{mk}} \eta_i U_i > 2p_{mk},\ m = k_0, ..., k-1\right), & \\ \qquad \times P\left(\sum_{j=1}^{p_{km}} \tau_j V_j \leqslant 2p_{km},\ m = k+1, ..., p\right), & k_0 \leqslant k \leqslant p, \end{cases}$$

where the U_i and V_j are all independent χ_1^2 random variables, the η_i and τ_j correspond to certain eigenvalues of Ω, and $p_{ij} \equiv (p-i)(q-i) - (p-j)(q-j)$.

Theorem 1 now follows immediately; when the population distribution is $(p+q)$-variate elliptical, the limiting covariance matrix of the vector z has the simple form

$$\Omega = (1+\kappa)\, I_{(p-k_0)(q-k_0)}.$$

Proof of Theorem 2. Assume that $K=k_0$ and put

$$f_k \equiv l_k + \frac{n}{2} \sum_{i=1}^{k_0} \log(1-\rho_i^2).$$

Note that l_k and f_k are maximized at the same value of k. Since r_i converges to ρ_i $(1 \leqslant i \leqslant p)$ in probability, we have

$$\frac{1}{n} f_k \to \frac{1}{2} \sum_{i=k+1}^{k_0} \log(1-\rho_i^2) \qquad (0 \leqslant k < k_0),$$

$$\frac{1}{n} f_k \to 0 \qquad (k_0 \leqslant k \leqslant p),$$

both in probability, as $n \to \infty$. This implies that

$$\begin{aligned} &\lim_{n \to \infty} P(\hat{K}_M = k) = 0, && (0 \leqslant k < k_0), \\ &\lim_{n \to \infty} P(\hat{K}_M = k) = \lim_{n \to \infty} P(f_k \geqslant f_m,\ m = k_0, ..., p) && (k_0 \leqslant k \leqslant p). \end{aligned} \tag{A.3}$$

Now, put

$$x_i = \frac{n^{1/2}(r_i^2 - \rho_i^2)}{2\rho_i(1-\rho_i^2)}, \qquad i = 1, ..., k_0,$$

$$x_j = n r_j^2, \qquad j = k_0 + 1, ..., p.$$

Then (see Hsu, 1941) $x_1, ..., x_{k_0}$ are asymptotically independent, and asymptotically independent of the set $(x_{k_0+1}, ..., x_p)$. The limiting distribution of x_i is standard normal, $i = 1, ..., k_0$, and the limiting distribution of $(x_{k_0+1}, ..., x_p)$ is the same as the distribution of the eigenvalues of a $(p-k_0)$ by $(p-k_0)$ matrix having the $W_{p-k_0}(q-k_0, I_{p-k_0})$ Wishart distribution. We can express f_k $(k_0 \leqslant k \leqslant p)$ in terms of the x's as follows:

$$\begin{aligned} f_k = {} & \frac{1}{2} \sum_{i=k_0+1}^{k} \Bigg\{ x_i + (k-q)\log(x_i) - 2 \sum_{\substack{j=k_0+1 \\ j>i}}^{k} \log(x_i - x_j) \\ & - \sum_{j=k+1}^{p} \log(x_i - x_j) \Bigg\} \\ & + \frac{1}{2} \sum_{i=1}^{k_0} \Bigg\{ (p+q+1)\log(1-\rho_i^2) + 2n^{1/2}\rho_i x_i + 2\rho_i^2 x_i^2 \\ & + (2k_0 - p - q)\log(\rho_i^2) - 2 \sum_{\substack{j=1 \\ j>i}}^{k_0} \log(\rho_i^2 - \rho_j^2) \Bigg\} \\ & - \frac{1}{2}\log(n)\{k_0(p+q-k_0-1)\} + \frac{k(p+q-k-1)}{2}\log(2) \\ & + \sum_{i=1}^{k} \left\{ \log\Gamma\left(\frac{p-i+1}{2}\right) + \log\Gamma\left(\frac{q-i+1}{2}\right) - \log(2\pi) \right\} + O_p(n^{-1/2}). \end{aligned} \tag{A.4}$$

The proof is now completed by substituting (A.4) into (A.3).

REFERENCES

Fujikoshi, Y. (1985). Two methods for estimation of dimensionality in canonical correlation analysis and the multivariate linear model. In *Statistical Theory and Data Analysis* (K. Matusita, Ed.), pp. 233–240. Elsevier Science, Amsterdam.

Fujikoshi, Y., and Veitch, L. G. (1979). Estimation of dimensionality in canonical correlation analysis. *Biometrika* **66** 345–351.

Glynn, W. J., and Muirhead, R. J. (1978). Inference in canonical correlation analysis. *J. Multivariate Anal.* **8** 468–478.

Gunderson, B. K. (1989). *Estimating the Dimensionality in Canonical Correlation Analysis.* Ph.D. dissertation. University of Michigan.

Hsu, P. L. (1941). On the limiting distribution of the canonical correlations. *Biometrika* **32** 38–45.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

Muirhead, R. J., and Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika* **67** 31–43.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist*. **6** 461–464.